Metrics for Openness

David Nichols

Department of Computer Science
University of Waikato
New Zealand
Outline

• Citation-based metrics
• altmetrics
• Open Access, Open Science
• Metrics to characterise Openness
Citation-based Metrics

• Journal Impact Factors,

• Total citations, avg. Citations
  – half-life, immediacy, EigenFactor

• h-index since 2005
  – Variants: g-index etc.

• Frequently used to characterise
  – Individuals, institutions, publication venues etc.
University of North Carolina Chapel Hill
North America
Publications: 0.628 | Citation Count: 152,932 | H-Index: 149
Top Areas: Psychiatry & Psychology, Pharmacology, Neuroscience, Diseases, Oncology

Duke University
North America
Publications: 0.812 | Citation Count: 164,954 | H-Index: 161
Top Areas: Neuroscience, Psychiatry & Psychology, Pharmacology, Cardiology, Oncology

Academic Ranking of World Universities in Computer Science - 2011

<table>
<thead>
<tr>
<th>World Rank</th>
<th>Institution</th>
<th>Country</th>
<th>Total Score</th>
<th>Score on HiCi</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>University of North Carolina at Chapel Hill</td>
<td>USA</td>
<td>47.9</td>
<td>0.0</td>
</tr>
<tr>
<td>36</td>
<td>Duke University</td>
<td>USA</td>
<td>47.4</td>
<td>35.4</td>
</tr>
</tbody>
</table>

Rank | Title                                    | Country    | Academic | Employer | Citations | Score |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>51-100</td>
<td>Duke University</td>
<td>United States</td>
<td>24.9</td>
<td>43.3</td>
<td>39.7</td>
<td></td>
</tr>
<tr>
<td>51-100</td>
<td>University of North Carolina, Chapel Hill</td>
<td>United States</td>
<td>17.8</td>
<td>24.0</td>
<td>57.3</td>
<td></td>
</tr>
</tbody>
</table>

http://academic.research.microsoft.com/Comparison?entitytype=7&id1=719&id2=814&topdomainid=2&subdomainid=0
• Citations are not enough
• Blogs, twitter, Mendeley, CiteULike etc.
• Article-level metrics

http://total-impact.org/collection/MqAnvl
Gudmundur Thorisson

@gThorisson
United Kingdom
Academic researcher based in Leicester, UK and Reykjavik, Iceland. Interested in applications of Internet identity and Linked Data in research data sharing.
http://gthorisson.name

Altmetrics
How often works by this author have been read, bookmarked or blogged by others

<table>
<thead>
<tr>
<th>Platform</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>citeulike</td>
<td>118</td>
</tr>
<tr>
<td>Mendeley</td>
<td>155</td>
</tr>
<tr>
<td>altmetric</td>
<td>10</td>
</tr>
</tbody>
</table>

Citations
How often works by this author have been cited by others

<table>
<thead>
<tr>
<th>Platform</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed Central</td>
<td>997</td>
</tr>
<tr>
<td>Academic Search</td>
<td>1468</td>
</tr>
<tr>
<td>Scopus</td>
<td>4380</td>
</tr>
<tr>
<td>CrossRef</td>
<td>2659</td>
</tr>
</tbody>
</table>

an altmetrics project.

http://sciencecard.org/gthorisson
• We start to confuse the “kind of use we can track” with use and “citation impact” with impact
  — Priem (2012)

• Also, we confuse the “act of publishing” with effective dissemination

• We measure the publication count, venue quality and citations but not other attributes of publications
Open Access

“By ‘open access’ to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.”

http://www.soros.org/openaccess/read

• Research Works Act, Elsevier boycott

Fake Elsevier @FakeElsevier
We've increased the freq. with which we say the words "Open Access". This is what you wanted, right? Everything can go back to normal now.
Results for #icanhazpdf

physiology  Michelle
#ICanHazPDF? tandfonline.com/doi/abs/10.120... 8 hours ago

LizNeeley  LizNeeley
#icanhazpdf please? Kasperon 1986, "Six Propositions on Public Participation and Their Relevance for Risk Communication" Risk Analysis 9 hours ago

rmounce  Ross Mounce
#icanhazpdf dx.doi.org/10.1063/1.3399... AIP paper. no access :( to ross.mounce @ gmail.com pls, thx!

kzelnio  Kevin Zelnio
"I never thought much about it until I couldn’t get a pdf copy of my own publication." Voices of #OpenAccess 1 Jan

OpenAccessHulk  OpenAccessHulk
OA HULK SMASH THOUGHTLESS IR MGRS! ALSO OPAQUE IR SOFTWARE! melissaterras.blogspot.com/2011/11/on-thu... 2 Nov
Scientists are urged to oppose new US legislation that will put studies behind a pay wall

BMJ 2012; 344: e452 (Published 17 January 2012)

Access to the full text of this article requires a subscription or payment. Please log in or subscribe below.
Missing Metrics

• We don’t have metrics that characterise authors’ behaviour with respect to Open Access issues
• So we can’t easily compare, chide or celebrate authors’ OAness

Villavelius Jan Velterop
Unfortunately it's 'publish or perish' and not 'share or be shunned' in the scientific ego-system. #SOPA #RWA

• Let’s fix that...
• “The University of North Carolina is dedicated to the *transmission* and advancement of knowledge and understanding”
  – Chapter 6, Section 600, Clause I, *The Code Of The Board of Governors, The University of North Carolina*
    • http://www.northcarolina.edu/policy/index.php?tag=100.1

• OA is part of “transmission”

• Not being measured because of a lack of a culture, and metrics, of measuring Openness
“What makes it complicated for students is that they don't understand why we would write articles, give them away, and be okay with it showing up on the Internet with a $35 price tag in the first place. Why make it so hard to get scholarly articles when sharing your research findings is the whole point?”
Openness Index

• Of the items you have published how many are free for anyone to read?
  – Practically: if someone searches for the title of your paper in Google do they find a copy they can read?
How to measure Openness?

• Grain size:
  – Person, Group, Dept, School, Institution, System
    • OA mandates (e.g. Harvard) will be evaluated at the institution/sub-institution level
  – Journal/Conference/etc, Publisher
  – Sub-discipline, Discipline
  – Funder
    • NSF Data Management Plans
  – State/Province, Country
“Björk et al. (2009) studied the annual volume and open access availability of scientific journal publishing. They estimated that after one year 11.3% of the scientific output in 2006 could be found in specific or institutional repositories or on the home pages of the authors.”
## Cross-Discipline

Norris et al. (OIR 2008)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Total articles</th>
<th>Percentage OA</th>
<th>Percentage TA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecology</td>
<td>628</td>
<td>34.39</td>
<td>65.61</td>
</tr>
<tr>
<td>Economics</td>
<td>966</td>
<td>54.45</td>
<td>45.44</td>
</tr>
<tr>
<td>Sociology</td>
<td>925</td>
<td>24.32</td>
<td>75.68</td>
</tr>
</tbody>
</table>

TA = Toll Access
“academic librarians ... OA choices compare favorably to the other authors in LIS journals”

Mercer (C&RL 2011)
Fig 1 Subscription based journal articles locatable with Google at non-journal websites, with approximate impact factors for 2002 in parentheses. No articles were found in Medline for *J Spinal Disord* from 2002-3

“Locatable with Google” – not necessarily legal copies

Wren (BMJ 2005)
Institution

• University of Helsinki
• 7771 journal articles from 2007-8
• 5% sample

Figure 1: Collective Open Access availability.

Koskinen et al. (SI 2010)
### Sub-Institution

#### Faculty (% of the sample)

<table>
<thead>
<tr>
<th>Faculty</th>
<th>Google</th>
<th>Google Scholar</th>
<th>HELDA</th>
<th>Open DOAR</th>
<th>Scientific Commons</th>
<th>Collective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty of Biological and Environmental Sciences (4.9%)</td>
<td>40.0%</td>
<td>65.0%</td>
<td>0.0%</td>
<td>5.0%</td>
<td>10.0%</td>
<td>70.0%</td>
</tr>
<tr>
<td>Faculty of Veterinary Medicine (3.4%)</td>
<td>35.7%</td>
<td>35.7%</td>
<td>7.1%</td>
<td>35.7%</td>
<td>28.6%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Independent institutes (10.3%)</td>
<td>52.4%</td>
<td>42.9%</td>
<td>2.4%</td>
<td>21.4%</td>
<td>23.8%</td>
<td>59.5%</td>
</tr>
<tr>
<td>Faculty of Pharmacy (2.7%)</td>
<td>9.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Faculty of Arts (2.5%)</td>
<td>40.0%</td>
<td>20.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>10.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Faculty of Behavioural Sciences (5.2%)</td>
<td>38.1%</td>
<td>28.6%</td>
<td>0.0%</td>
<td>14.3%</td>
<td>9.5%</td>
<td>42.9%</td>
</tr>
<tr>
<td>Faculty of Medicine (47.2%)</td>
<td>44.8%</td>
<td>41.7%</td>
<td>0.5%</td>
<td>13.5%</td>
<td>13.5%</td>
<td>49.5%</td>
</tr>
<tr>
<td>Faculty of Agriculture and Forestry (6.9%)</td>
<td>25.0%</td>
<td>32.1%</td>
<td>0.0%</td>
<td>3.6%</td>
<td>7.1%</td>
<td>35.7%</td>
</tr>
<tr>
<td>Faculty of Science (15.2%)</td>
<td>50.0%</td>
<td>38.7%</td>
<td>0.0%</td>
<td>21.0%</td>
<td>24.2%</td>
<td>56.5%</td>
</tr>
<tr>
<td>Faculty of Law (1.2%)</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Faculty of Theology (1.0%)</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Faculty of Social Sciences (6.1%)</td>
<td>56.0%</td>
<td>40.0%</td>
<td>0.0%</td>
<td>16.0%</td>
<td>16.0%</td>
<td>60.0%</td>
</tr>
</tbody>
</table>

*Found*

**Figure 3:** Open Access availability by faculty.

*The sum is more than 100% because of the co-authored articles across faculties. (Figure 3)*

Faculties of the University of Helsinki

Koskinen et al. (SI 2010)
Repository Statistics

Key Figures

<table>
<thead>
<tr>
<th>Collection</th>
<th>Total Usage</th>
<th>Yesterday</th>
</tr>
</thead>
<tbody>
<tr>
<td>35,265 Works</td>
<td>6,273,655 Downloads</td>
<td>3,691 Downloads</td>
</tr>
<tr>
<td>55.2% Full-text</td>
<td>97.5% External</td>
<td>4 Deposits</td>
</tr>
<tr>
<td>48.7% Open-access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Repository - Individual

Statistics Overview
Skitmore, Martin

Key Figures

<table>
<thead>
<tr>
<th>Collection</th>
<th>Total Usage</th>
<th>Last 28 Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>214 Works</td>
<td>123,148 Downloads</td>
<td>2,134 Downloads</td>
</tr>
<tr>
<td>98.1% Full-text</td>
<td>98.9% External</td>
<td></td>
</tr>
<tr>
<td>94.9% Open-access</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individual Level

• Missing?

• Just like the h-index for citations

• Each author gets an Openness Index defined across their publications
Openness Index

• = Openly Accessible Publications / Total Publications

• Example:
  – Time period: 2011
  – Publication type:
    • conference papers (C), journal papers (J)

\[ OI_{2011}^{C,J} = 0..1 \]
iSchool Dean *Openness Index 2011*

1  Georgia Tech, Illinois, IU-SIC, Syracuse, Toronto, UCD, UC-Irvine, UCL, UMD, UMich, UNC, UW
0.5  Tampere, UBC
0.25  CMU
0.2  Humboldt
0  IU-SLIS, PSU, RSLIS, Sheffield, UK, UMBC

No data: Drexel, FSU, Pitt, Rutgers, SMU, UC-Berkeley, UCLA, UNT, UT-Austin
No evaluation: Nanjing, Wuhan
Visualising Individual Openness

OI = 0.7

2011

2012-02-19

What to measure?

• Just proportion?
  – Should 1 closed access paper get the same value as 10 closed access papers?
  – Do pre-peer review papers count as anything?
    • Wiley-Blackwell only allow pre-review versions to be archived in an IR

• OA-gratis v OA-libre
  – Same value?

• Publication OA status
  – Equal value for an OA journal v self-archive?

• Publication type
  – Is C + J a reasonable middle ground?
  – Book chapters, books, ...

• Location
  – Archive-friendly v (temporary) faculty web space
OA-libre v OA-gratis

• OA-gratis
  – Free to read

• OA-libre (CC-BY)
  – OA-gratis
  – +
  – Free to text-mine
  – Free to remix
  – etc.
Preservation

• “Knowing that faculty Web sites are deleted after they leave the university and that the maintenance of departmental servers varies over time, some faculty interviewed in 2006 expressed concern about the preservation of their ‘legacy’” – Covey (2009)

• Simple OI doesn't take into account the location
  – personal web space as good as IR?
Legacy-Friendly OI

• Yes: open access journal, publisher DL (e.g. HICSS at IEEE), IR
• No: personal web/FTP space
• ? Research group web space
Effective Openness Index

• After taking account of existing copyright agreements

• What proportion of the research items you could have made open – have you actually made open?
  – Difficult to automate
  – Individual author addenda, special issue one-off copyright arrangements
Open Acce$$$ Cost Index

• How much does it cost to read your research?

• For all the items that are not open
  – Sum the cost to access them
  – Simple: individual independent items
  – Complex: bulk deals, joining societies etc

• At the institution level – what is UNC’s OACI?
  – State legislature “pays twice”
  – What would it cost v what does it actually cost the public?
9162

IN ASSEMBLY

January 31, 2012

Introduced by M. of A. HEVESI -- read once and referred to the Committee on Governmental Operations

AN ACT to establish the taxpayer access to publicly funded research act

THE PEOPLE OF THE STATE OF NEW YORK, REPRESENTED IN SENATE AND ASSEMBLY, DO ENACT AS FOLLOWS:

Section 1. Short title. This act shall be known and may be cited as the "taxpayer access to publicly funded research act".

Section 2. (a) As used in this act, the term:

(i) "Agency" shall mean any subdivision of the state, including, but not limited to any agency, commission or authority;

(ii) "Direct research" shall mean research resulting directly from state agency funding; and

(iii) "Policy" shall mean the public access policy established pursuant to subdivision (b) of this section.
The Rental Market for Research

http://www.deepdyve.com/
Open Science

• “Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process.”
  – Nielsen (2011)

• Data
• Code

http://www.openscience.org/blog/?p=454
The (in)famous Jones quote

“Even if WMO [World Meteorological Organization] agrees, I will still not pass on the data. We have 25 or so years invested in the work. Why should I make the data available to you, when your aim is to try and find something wrong with it?” 21 Feb 2005

Professor Phil Jones
Director of the Climatic Research Unit (University of East Anglia, UK) to “amateur climate researcher” (and sceptic) Warwick Hughes about CRU weather station data

http://en.wikipedia.org/wiki/Phil_Jones
http://www.climateaudit.org/?p=403
Data Archiving Index

- “Lonnie Thompson is one of the worst archiving offenders in paleoclimate, and that’s a real beauty contest.”

http://climateaudit.org/2006/07/06/new-thompson-article-at-pnas/#comment-55284

- This statement cannot be evaluated with current metadata – but it implies the existence of paper and author-specific data archiving metadata – and a ranking of researchers ordered by their data archiving rates
Data Archiving Index

• If a paper creates/uses a dataset does it
  – Archive or uniquely identify the data in an open location

• = papers with open data / papers with data

• Lots of data release issues to get a ‘fair’ index
  – Privacy, commercial, ...
Code Archiving Index

• Repeat Data Index but for source code

• “anything less than the release of source programs is intolerable for results that depend on computation”
  – Ince, Hatton & Graham-Cumming (Nature 2012)
Faculty & Copyright

• Unaware, confused, ignoring

• “Many faculty members are simply not aware of publisher policies. Many have a meager understanding of copyright. Some faculty have little respect or concern for publisher policy or copyright.” – Covey (2006)
• “many authors are not aware of the terms of their copyright agreements and many believe they hold copyright to their own works. At the same time, authors who know their rights or publisher requirements do not necessarily abide by them”
  – Antelman (2006)
Illegality Index

• How much of your work is available in *contravention* of existing copyright agreements?
  
  • = num. illegally available / num. of items

• At the Repository level this is a measure of:
  – Workflow copyright checking effectiveness
  – Faculty mistakes
Reversion Index

- Open access can disappear
  - Change in OI, but worth separating?

Take Down Policy

November 2009

There exists the possibility that complaints arise regarding alleged copyright or licensing infringement violations pertaining to material included in the Carolina Digital Repository (CDR), or material that is deemed to be potentially libelous, plagiarized, or legally offensive. In the event that such a complaint is made, it will be referred to the Repository Librarian with one working day of receipt. The Repository Librarian will be responsible for assessing, within a second working day, what risks may be incurred by leaving the offending materials accessible or visible. If such risk seems likely, the Repository Librarian will take action to make the digital object inaccessible, although some metadata relating to the material may remain searchable and viewable.

The Repository Librarian should inform the depositor of the material that a complaint has been lodged. The Librarian may also seek the advice of University Counsel in determining the ultimate resolution of the complaint. A record of the complaint and its resolution should be recorded with the metadata for the collection of which the object or objects are a part and associated with the submission agreement under which the material was acquired.

https://cdr.lib.unc.edu/external?page=about.takeDownPolicy
Clayton Weatherston's work erased

NEIL REID

Last updated 05:00 11/12/2011

Red-faced Otago University officials have deleted killer Clayton Weatherston's work from an online archive after offending the family of murder victim Sophie Elliott.

The university spiked seven papers written by Weatherston between 2001 and 2006 after being approached by the Sunday Star-Times last week.

Weatherston completed a PhD at the university before tutoring in economics. But in 2008 the 36-year-old killed Elliott, 22, stabbing her more than 200 times in a bedroom of her family’s Dunedin home.

Weatherston, who had taught Elliott before they embarked on a five-month relationship, was later jailed for a minimum non-parole period of 18 years.

After the case, university officials tried to erase references to Weatherston from its records. But the seven items by the convicted murderer, covering professionalism in rugby, crowd sizes in Super Rugby, and evaluating rugby’s former NPC, slipped through the cracks.

A university spokeswoman did not know Weatherston’s work was still online when contacted by the Sunday Star-Times, but rang back later, saying references to Weatherston had been removed.

http://www.stuff.co.nz/national/crime/6121173/Clayton-Weatherstons-work-erased
(Missing) Openness Metrics

- Openness Index
  - Effective Index: given past copyright
- Openness Cost Index
- Illegality Index
- Data Archiving Index
- Code Archiving Index
- Reversion Index
Value of (yet) more metrics?

• “What gets measured gets noticed”
• “What gets measured gets managed” - Drucker
• “What gets measured gets to frame the issue”

• “The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”
  • http://en.wikipedia.org/wiki/Campbell%27s_law
    – Citation cartels, coercive citation etc
  • El Naschie, *Chaos, Solitons & Fractals*
  • http://elnaschiewatch.blogspot.com/

• Can you subvert an altmetrics measure?
• Can you subvert an Openness Index?